

Article

DOA Estimation Using Deep Neural Network with Angular Sliding Window

Yang Li ^{1,2,3,4}, Zhanhu Huang ^{1,3,4}, Can Liang ^{1,3,4} , Liang Zhang ^{1,3} , Yanhua Wang ^{1,2,3,4,5,*}, Junfu Wang ⁶, Yi Zhang ⁶ and Hongfen Lv ⁶

¹ Radar Research Laboratory, School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

² Beijing Institute of Technology Chongqing Innovation Center, Chongqing 401120, China

³ Electromagnetic Sensing Research Center of CEMEE State Key Laboratory, School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

⁴ Beijing Key Laboratory of Embedded Real-Time Information Processing Technology, Beijing 100081, China

⁵ Advanced Technology Research Institute, Beijing Institute of Technology, Jinan 250300, China

⁶ Beijing Racobit Electronic Information Technology Co., Ltd., Beijing 100081, China

* Correspondence: wyhlucky@bit.edu.cn

Abstract: Deep neural network (DNN) has shown great potential in direction-of-arrival (DOA) estimation. In high dynamic signal-to-noise (SNR) scenarios, the estimation accuracy of the weaker sources may degrade significantly due to insufficient training samples. This paper proposes a deep neural network framework with sliding window operation. The whole field-of-view (FOV) is divided into a series of sub-regions via sliding windows. Each sub-region is assumed to contain one source at most. Thus, the single-source data can be used to train all the networks, alleviating the need for the training samples and the prior information on the number of sources. A detector network and an estimator network are followed for each sub-region, enabling high estimation accuracy and the number of sources. Simulation and real data experiment results show that the proposed method can achieve excellent DOA and source number estimation performance. Specifically, in the real data experiment, the results show that the RMSE of the proposed method reaches 0.071, which is at least 0.03 lower than FFT, MUSIC, ESPRIT, and a deep learning method namely deep convolutional network (DCN), cannot estimate the lower SNR source in high dynamic SNR scenarios.

Keywords: array signal processing; direction-of-arrival (DOA) estimation; deep neural network (DNN); supervised learning



Citation: Li, Y.; Huang, Z.; Liang, C.; Zhang, L.; Wang, Y.; Wang, J.; Zhang, Y.; Lv, H. DOA Estimation Using Deep Neural Network with Angular Sliding Window. *Electronics* **2023**, *12*, 824. <https://doi.org/10.3390/electronics12040824>

Academic Editors: Yin Zhang, Yulin Huang, Yachao Li, Deqing Mao and Maciej Lawryńczuk

Received: 11 January 2023

Revised: 29 January 2023

Accepted: 3 February 2023

Published: 6 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Direction-of-arrival (DOA) estimation is an essential task in array signal processing due to the extensive applications in radar, wireless communications, and acoustics [1]. Model-based DOA estimation approaches have been comprehensively investigated, including Fourier Transform (FT), subspace methods, and compressed sensing (CS) [2]. FT is easy to implement with high computational efficiency and robustness [3]. However, its angular resolution is restricted by the size of the array. Subspace methods, such as multiple signal classification (MUSIC) [4] and estimation of signal parameters via rotational invariance techniques (ESPRIT) [5], can break the resolution barrier. They utilize the orthogonality between the signal subspace and the noise subspace, which usually requires multiple snapshots to estimate the covariance matrix and the underlying subspaces [6]. CS methods solve the angular observation equation by introducing certain sparse regularizations such as L_0 or L_1 -norm [7,8]. The accuracy and resolution are improved at a much higher computational cost [9].

Despite the wide applications, model-based approaches are susceptible to adverse conditions, such as low signal-to-noise ratio (SNR) and array imperfections. Recently,

owing to the powerful nonlinear fitting capabilities, deep neural network (DNN) has been introduced to DOA estimation [10,11]. As a data-driven approach, DNN can deal with adverse situations given sufficient training samples. The key concept of existing DNN methods is discretizing the field of view (FOV) and transforming DOA estimation into a multi-label classification task. To cope with the low SNR conditions, the networks are trained across a range of low SNRs and outperform their competitors in the low SNR regime [12,13]. To obtain good adaptation to array imperfections, autoencoders are constructed to reduce the influence of noise and array imperfections, after which the network performs satisfyingly in imperfection adaptation [14]. In addition, a new deep learning architecture is proposed based on the imperfect array, and the output of the network is a vector for the spectrum estimation, which can avoid discretizing the spatial domain [15]. Furthermore, to deal with the grid mismatch, some networks are proposed to exploit the Toeplitz property and reconstruct the covariance matrix, and model-based approaches, such as MUSIC or root-MUSIC, are applied to obtain gridless DOA [16,17].

DNN has shown great potential in adverse conditions. However, high dynamic SNR is still a limiting factor in the accuracy of DOA estimation. DNN requires enough training samples that cover all the practical situations. When the number of sources is large, it is difficult to cover all possible combinations of positions and amplitudes. When the SNRs of the sources vary significantly, weaker sources achieve much lower accuracy or may even be missed. Moreover, the number of sources is usually needed as the prior information. The number of the resulting training samples is huge, which leads to very high computational costs as well [13].

In this paper, we propose a DNN framework with a sliding window (DNN-SW) for DOA estimation to cope with the high dynamic SNR scenarios. The entire FOV is divided into a series of overlapping angular sub-regions through the sliding window operation. The core network consists of a detector network and an estimator network. The detector determines whether the sub-region contains a source, and the estimator informs its angle. This paper assumes that there is only one source in each sub-region. Based on the assumption, single-source data is used to train the networks, and the requirement of the training sample is alleviated. Compared with existing methods based on DNN, the DOA estimation task is jointly accomplished by multiple networks, and each network only estimates the angle of the current sub-region, greatly simplifying the task of each network and improving network performance in high dynamic scenarios. Furthermore, the number of sources can be obtained adaptively according to the number of sub-regions in which the source exists, which adapts to the number of sources.

The rest of the paper is organized as follows. In Section 2, we present the signal model. In Section 3, we introduce the input data preprocessing and network structure in detail, including the sliding window module, detection module, DOA estimation module, and angle merging. In Section 4, we offer the results of the simulation to prove the advantageous performance of the proposed method by comparing the other approaches and collecting the real data for verification. Finally, conclusions are summarized in Section 5.

2. Signal Model

In this paper, the single snapshot scenario is considered [18]. As shown in Figure 1, we consider a uniform line array (ULA) with N elements in the narrow-band mode. It is assumed that there are K far-field sources from $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_K]^T$, θ_k indicates the direction of the k th source, and the received signal can be formulated as

$$\mathbf{x} = \sum_{k=1}^K \mathbf{a}(\theta_k) S_k + \mathbf{n} = \mathbf{A}(\boldsymbol{\theta}) \mathbf{S} + \mathbf{n}. \quad (1)$$

where $\mathbf{S} = [S_1, S_2, \dots, S_K]^T$ represents the amplitude of the sources and $\mathbf{n}(t)$ is a $N \times 1$ vector denoting the statistically independent white Gaussian noise vector with zero mean value and unknown variance σ_n^2 . The steering vector $\mathbf{a}(\theta_k)$ can be expressed as

$$\mathbf{a}(\theta_k) = \left[1, e^{j\frac{2\pi}{\lambda}d \sin(\theta_k)}, \dots, e^{j\frac{2\pi}{\lambda}(N-1)d \sin(\theta_k)} \right]^T. \tag{2}$$

where $\lambda = c/f$ is the wavelength of the transmitted signal, f is the carrier frequency and c is the speed of the light. Additionally, d describes the array element space. As for $\mathbf{A}(\boldsymbol{\theta})$, it is the array manifold matrix, which can be written as

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \exp\left(j2\pi\frac{d}{\lambda} \sin \theta_1\right) & \exp\left(j2\pi\frac{d}{\lambda} \sin \theta_2\right) & \dots & \exp\left(j2\pi\frac{d}{\lambda} \sin \theta_K\right) \\ \vdots & \vdots & \ddots & \vdots \\ \exp\left(j2\pi\frac{(N-1)d}{\lambda} \sin \theta_1\right) & \exp\left(j2\pi\frac{(N-1)d}{\lambda} \sin \theta_2\right) & \dots & \exp\left(j2\pi\frac{(N-1)d}{\lambda} \sin \theta_K\right) \end{bmatrix} \tag{3}$$

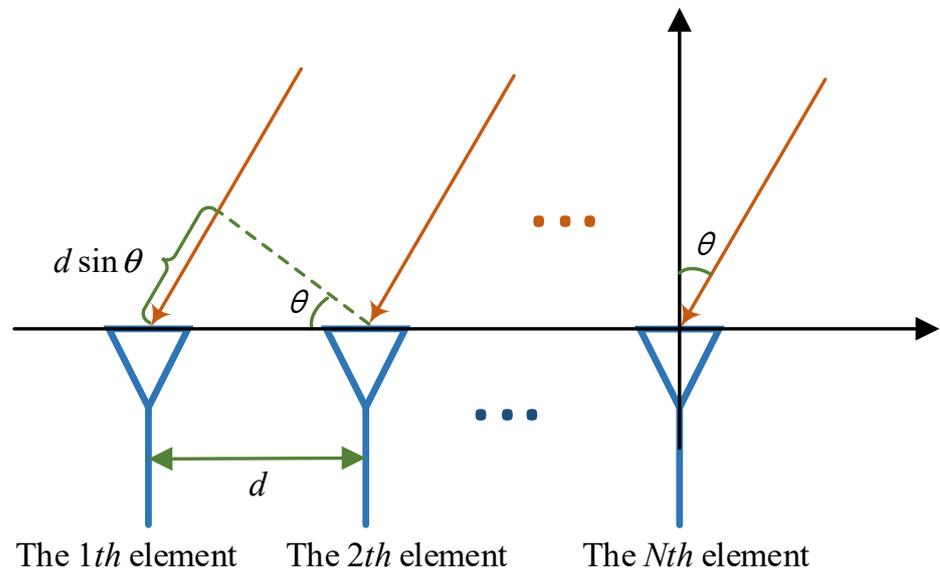


Figure 1. The system model for the DOA estimation with the far-field narrowband source.

3. The Proposed Method

The structure of DNN-SW is shown in Figure 2. it is divided into several overlapping sub-regions and assume the sources are located in different sub-regions independently. Specifically, there are four main modules in the proposed network. First, in the sliding window module, the FOV is split into a series of overlapping sub-regions in which there is one source at most. Then, the detection module and DOA estimation module are followed. The detection module contains multiple detectors for each sub-region. The detector network determines whether a source exists in the corresponding sub-region, which is formed as a binary classification task. Similar to the detection module, the estimation module also includes multiple estimator networks to obtain the angles of the sources. Since the sub-regions are overlapped, one source can be detected and estimated in several sub-regions. Therefore, an angle merging module is applied to give the final results.

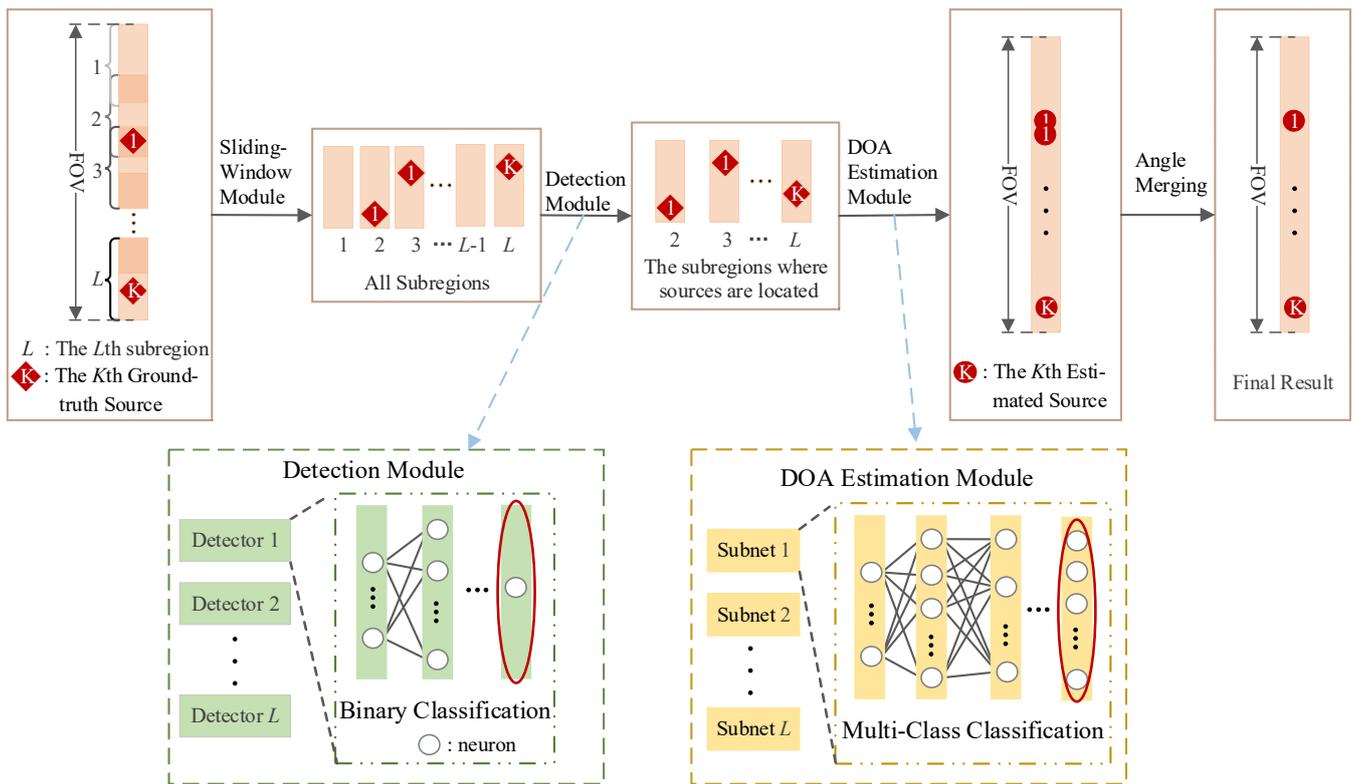


Figure 2. The architecture of DNN-SW for DOA estimation.

3.1. Input Data Preprocessing

To retain the amplitude and phase information of the single snapshot signal \mathbf{x} [19], we extract four parts of information from \mathbf{x} , including the real part, imaginary part, angle value, and modulus. The input vector \mathbf{X} can be described as

$$\mathbf{X} = [\text{Re}(\mathbf{x}^T) \quad \text{Im}(\mathbf{x}^T) \quad \angle(\mathbf{x}^T) \quad \text{abs}(\mathbf{x}^T)]. \quad (4)$$

3.2. Sliding Window Module

To alleviate the negative influence of different sources in the high dynamic SNR scenario, the sliding window module is proposed. By detecting and estimating the sources separately through the network, the information of each source can be focused on and easily extracted. Furthermore, based on this structure, the detectors and estimators are trained using single-source data. Compared to other methods [11–14], we can greatly reduce the number of training samples and the training time cost. Here, the range of the FOV is $[\varphi_{\min}, \varphi_{\max})$, it is divided into several overlapping sub-regions, as shown in Figure 3. In this paper, the range of the sub-region δ_1 is set to be close to $\Delta\theta_{3dB}$ (the 3 dB beamwidth of the array), resulting in the situation that at most one source can be estimated in each sub-region. According to [20], $\Delta\theta_{3dB}$ can be calculated by

$$\Delta\theta_{3dB} = \frac{0.886\lambda}{d(N-1)\cos(\theta_m)} \quad (5)$$

where, θ_m is the center of the beamwidth. Moreover, δ_2 is the step size when dividing the FOV. To alleviate the missing problem of sources and improve the estimation accuracy, each source is set in multiple sub-regions for repeated detection and estimation, resulting in $\delta_2 \leq \delta_1/2$. Thus, in this paper, δ_2 is configured to be $\delta_1/2$, and the number of sub-regions L is

$$L = \left\lceil \frac{\varphi_{\max} - \varphi_{\min} - \delta_1}{\delta_2} + 1 \right\rceil. \quad (6)$$

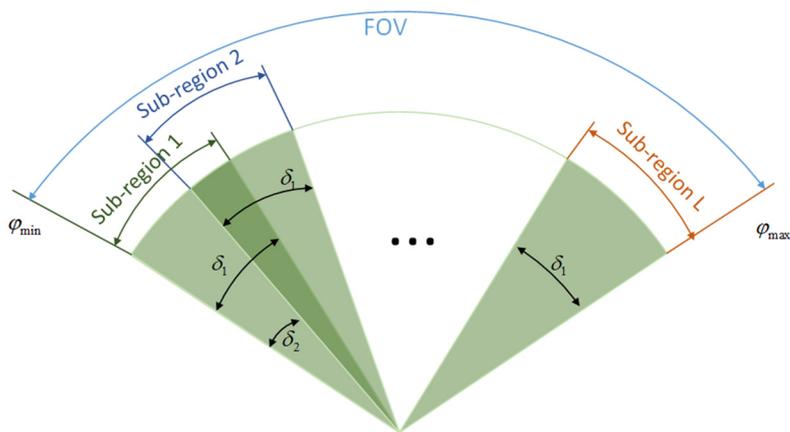


Figure 3. The subregions division method.

3.3. Detection Module

The structure of this module is shown in Figure 2. L detector networks are constructed to accomplish the detection task in L sub-regions. The data is fed into the L detectors to decide which sub-regions the sources are located in. Since this is a supervised task, the training and testing details are introduced, respectively. During the training process, take the i th detector as an example; given the j th training data \mathbf{X}_j^{train} , the output \tilde{d}_{ij} can be obtained

$$\tilde{d}_{ij} = f_W(f_{W-1}(\dots f_1(\mathbf{X}_j^{train}))), \tag{7}$$

where f_1 to f_W are W fully connected layers. f_1 to f_{W-1} are all followed by a rectified linear unit (ReLU) layer. Finally, the Tanh layer is applied to generate the final detection. The ReLU layer and Tanh layer are defined as

$$\text{ReLU}(x) = \max(0, x) \tag{8}$$

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{9}$$

Furthermore, d_{ij} represents the ground-truth label of the sample \mathbf{X}_j^{train} . If \mathbf{X}_j^{train} is in the i th sub-region, the label d_{ij} is set to 1; otherwise, d_{ij} is set to 0. For each detector, the mean square error is served as the loss function for backpropagation to evaluate the detection performance. The objective function can be written as

$$\text{loss}_1 = \frac{1}{J} \sum_{j=1}^J (d_{ij} - \tilde{d}_{ij})^2, \tag{10}$$

where J denotes the number of all the training samples. The parameters of the network are updated by minimizing (10) through the adaptive moment estimation (Adam) optimizer.

In the testing process, a threshold Th_1 is designed to judge whether the source lies in i th sub-region. Its value is determined by the statistical analysis results of training samples to guarantee a satisfying detection. The details will be discussed in the simulation section. When a testing sample \mathbf{X}_j^{test} is put into the i th detector, y_{ij} is the detection result, which is defined as

$$y_{ij} = \begin{cases} 1, & \text{if } \tilde{d}_{ij} > Th_1 \\ 0, & \text{else} \end{cases} \tag{11}$$

When $y_{ij} = 1$, the source is considered in this sub-region, otherwise, it does not exist in this sub-region.

3.4. DOA Estimation Module

After detecting which sub-regions the sources are in, the DOA estimation module is built to achieve the specific angle estimation task. As shown in Figure 2, similar to the detection module, L estimator networks are established for the L sub-regions. For each estimator, the sub-region is further divided into a series of grids with the step of δ_3 and the number of grids M is $\lfloor \delta_1 / \delta_3 \rfloor$. During the training process, to avoid the negative effects of error detection, estimators are trained independently of the detectors. We take the i th estimator into account, given the sample \mathbf{X}_j^{train} , the input data of this estimator is $d_{ij} \times \mathbf{X}_j^{train}$, which means only the data in the i th sub-region is trained. In the estimating phase, the label is encoded by the one-hot encoding method, which is a vector that represents the probabilities of all the alternative angles. The ground-truth label \mathbf{P}_{ij} of \mathbf{X}_j^{train} for the i th estimator can be written as $\mathbf{P}_{ij} = [p_{ij1}, p_{ij2}, \dots, p_{ijM}]^T$. If the source locates in the m th grid, p_{ijm} is 1, otherwise, p_{ijm} is 0. When training, the estimation results of the estimator $\tilde{\mathbf{P}}_{ij}$ can be expressed as

$$\tilde{\mathbf{P}}_{ij} = f_Z \left(f_{Z-1} \left(\dots f_1 \left(d_{ij} \times \mathbf{X}_j^{train} \right) \right) \right) = [\tilde{p}_{ij1}, \tilde{p}_{ij2}, \dots, \tilde{p}_{ijM}]^T, \tag{12}$$

where f_1 to f_Z are Z fully connected layers. f_1 to f_{Z-1} are followed by a ReLU layer and f_Z is followed by a Softmax layer. The Softmax layer is defined as

$$\text{Softmax}(\tilde{p}_{ijm}) = \frac{\exp(\tilde{p}_{ijm})}{\sum_h \exp(\tilde{p}_{ijh})}. \tag{13}$$

According to the ground-truth label \mathbf{P}_{ij} and the predicted label $\tilde{\mathbf{P}}_{ij}$, cross-entropy is selected as the loss function

$$\text{loss}_2 = -\frac{1}{J} \sum_{j=1}^J \sum_{m=1}^M p_{ijm} \log(\tilde{p}_{ijm}), \tag{14}$$

where J denotes the number of training samples. The parameters of the estimator are updated by minimizing (14) through the Adam optimizer.

In the testing process, for the i th estimator, the sample \mathbf{X}_j^{test} is served as the input data based on the detection result. The input can be formulated as $y_{ij} \times \mathbf{X}_j^{test}$. Only when y_{ij} is 1, does the output of the estimator makes sense. The location with max probability is used to calculate the estimated angle $\tilde{\theta}_{ij}$, and it can be expressed as

$$\tilde{\theta}_{ij} = \text{argmax}(\tilde{\mathbf{P}}_{ij}) + (j - 1) \times \delta_2 + \varphi_{\min}. \tag{15}$$

3.5. Angle Merging

Through the detection and DOA estimation module, the source is first detected, and then the specific angle is obtained. However, although the adjacent overlapping sub-regions allow the detection results to be more accurate and complete, the same source will be detected and estimated repeatedly in multiple sub-regions. Namely, several angles may be obtained according to one source. An angle merging algorithm is proposed to solve the angle redundancy problem. The estimated angles from the same source are considered to have minor differences so that they can merge into one angle as the final output. The fusion threshold Th_2 is introduced to decide whether the angles should be fused. Since high-resolution DOA estimation is not considered in this paper, Th_2 is set to be a little lower than $\Delta\theta_{3dB}$. If the difference between the estimated angles $|\tilde{\theta}_l - \tilde{\theta}_h|$ is lower than Th_2 , they will be merged by

$$\tilde{\theta}' = \frac{1}{2} (\tilde{\theta}_l + \tilde{\theta}_h) \text{ if } \text{abs}(\tilde{\theta}_l - \tilde{\theta}_h) < Th_2. \tag{16}$$

After the merging algorithm, the final angles estimated are obtained, and the number of sources can be acquired automatically.

4. Experiment

In this section, we conduct the DOA estimation based on the simulation data and real data to evaluate the proposed method. First, the simulation data is used to verify the effectiveness and advantages of our method. Then, we collected real radar signals to assess the performance of DNN-SW in practical application.

4.1. Simulation Settings

In the simulations, the FOV is the range of $[-60^\circ, 60^\circ]$, and a 40-element uniform linear array with $\lambda/2$ inter-element spacing is considered. According to (5), since the $\Delta\theta_{3dB}$ of different beamwidth centers are different, in this paper, the center of FOV is set as the beamwidth center θ_m . When θ_m is 30° , $\Delta\theta_{3dB}$ is 3.01° . The range of sub-region δ_1 is set to 3° and the step size δ_2 is set to 1.5° . As a result, the sub-region number L is 79. Additionally, the sampling interval δ_3 is specified as 0.1° so that the grid number for each sub-region M is 30 and the grid number for FOV is 1200 categories of direction in total. The detailed parameters are listed in Table 1.

Table 1. Parameter Settings.

Parameter	Value
<i>Sensor array</i>	
Configuration	ULA
Inter-element spacing	$d = \lambda/2$
The number of the element	$N = 40$
<i>Sub-region</i>	
FOV	$[\varphi_{min}, \varphi_{max}] = [-60^\circ, 60^\circ]$
Range of sub-region	$\delta_1 = 3^\circ$
Step of sub-region	$\delta_2 = 1.5^\circ$
Interval of the FOV	$\delta_3 = 0.1^\circ$
Threshold of detector ¹	$Th_1 = 0.25$
Threshold of angle merging ¹	$Th_2 = 2^\circ$
<i>Detector and Estimator Network</i>	
Hidden layers, # Detector Network	16, 1
Hidden layers, # Estimator Network	128, 256, 30
Activation function >	Detector: Tanh; Estimator: Softmax

¹ The value of Th_1 and Th_2 are described in Section 4.3.

For the training dataset, the SNR of the source is 15 dB, and 30 samples are collected in each direction. Therefore, there are 36,000 training samples in total. For each detector and estimator, the size of training data is 36,000 and 900, respectively. For all the experiments, the SNR is defined in [13]

$$SNR = 10 \log_{10} \frac{\min(\sigma_1^2, \dots, \sigma_K^2)}{\sigma_n^2}, \quad (17)$$

where σ_i^2 represents the power of the i th source, $i = 1, \dots, K$. σ_n^2 represents the power of the noise.

For each detector, the number of neurons per layer of the network is {16, 1} with a batch size of 128 during 100 training epochs. Similarly, for each estimator, the number of neurons per layer of the network is {128, 256, 30} with a batch size of 128 during 200 training epochs. Moreover, the learning rate is configured to 0.001 for all networks.

The simulations are carried out in a workstation with MATLAB R2022a, Intel Xeon Gold 6240 processor at 2.60GHz, and NVIDIA A100 Tensor Core GPU. The detector net-

works and estimator networks are based on Pytorch 1.11.0 and Python 3.9.12. Based on the conditions in the training process, the average running time of all the 79 detection networks and estimator networks is about 90.2 s and 5.6 s, respectively. In the testing process, each detection network and each estimator network respectively cost 3.09 us and 3.89 us, which is obtained by calculating the average running time of 1000 testing samples.

4.2. Evaluation Metrics

In the simulations, to objectively and effectively evaluate the performance of the DNN-SW, two evaluation metrics are utilized, including Acc and root mean square error (RMSE). Since the number of sources is unknown, Acc is an important metric to evaluate DNN-SW. It describes the percentage of the number of testing samples whose source numbers are estimated correctly by the network [21]. It can be formulated as

$$Acc = \left(\frac{1}{m} \sum_{i=1}^m p_i \right) \times 100\%, \quad (18)$$

where

$$p_i = \begin{cases} 1, & \text{if } \text{num}(\tilde{\theta}_i) = \text{num}(\theta_i) \\ 0, & \text{others} \end{cases}. \quad (19)$$

Here, θ_i and $\tilde{\theta}_i$ denotes the ground truth and the prediction directions of the i th testing sample, respectively, $i = 1, 2, \dots, m$, where m denotes the number of testing samples.

Additionally, RMSE is also a classic and common metric in past research [13,22]. We calculate the RMSE of the testing samples whose source number is estimated correctly. RMSE can be obtained by

$$RMSE = \sqrt{\frac{1}{HQ} \sum_{h=1}^H \left[\sum_{q=1}^Q (\tilde{\theta}_{h,q} - \theta_{h,q})^2 \right]}, \quad (20)$$

where H represents the number of samples whose source number is estimated correctly, and Q represents the number of the source in a testing sample. $\theta_{h,q}$ and $\tilde{\theta}_{h,q}$ denote the q th estimated direction and ground-truth direction of the h th sample, respectively.

4.3. Determination of Th_1 and Th_2

In this part, the determination methods of Th_1 and Th_2 are described in detail. Since the detection process can be regarded as a binary task and Th_1 is an important threshold to decide the detection results, F_1 score is served as the criterion to select the optimal parameter Th_1 . In binary classification tasks, the F_1 score is widely used to analyze the accuracy of machine learning models [23–26]. It takes both the *precision* and *recall* of the model into account to provide an objective description of the method.

In order to obtain the F_1 score for the detector network, the samples can be split into four parts according to their ground truth and predicted labels, as shown in Table 2.

Table 2. The standard confusion matrix.

	Predicted Positive	Predicted Negative
Actual positive	TP	FN
Actual negative	FP	TN

In the detector network, the sample is considered positive if its source is in the corresponding sub-region. According to [25], *precision* and *recall* are first calculated by

$$precision = \frac{TP}{TP + FP}, \quad (21)$$

$$recall = \frac{TP}{TP + FN}, \quad (22)$$

where *precision* is the proportion of the positive predicted samples in the actual positive samples, and *recall* is the proportion of the actual positive samples in all predicted positive samples. It should be noted that since there are $L = 79$ detectors in our method, the number of samples used to obtain the *precision* and *recall* are the overall number of samples of all 79 detectors. In this case, the overall performance of the detection module is assessed, and the results will not be influenced by the extreme results of some detectors. Then, F_1 score is obtained by calculating the harmonic mean of *precision* and *recall*

$$F_1 = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (23)$$

where the range of F_1 is $[0, 1]$. If all the positive samples are wrongly predicted, F_1 is equal to 0 which is the minimum value. Additionally, when the samples are all correctly predicted, F_1 is equal to 1, which is the maximum value.

In the simulation, to select the best threshold Th_1 , we randomly generate 10,000 testing samples. The samples contain two sources, which are not located in one sub-region, and the SNR of these sources is configured to 15 dB. According to (23), we calculate the F_1 for each Th_1 with the step of 0.05 from 0 to 1, and the results are shown in Figure 4.

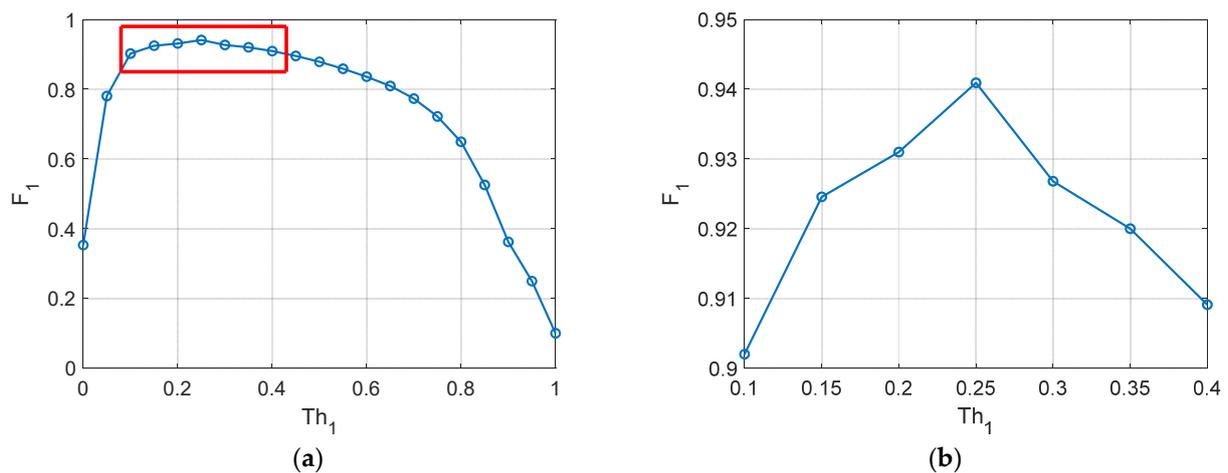


Figure 4. (a) The value of F_1 for each Th_1 with the step of 0.05 from 0 to 1. (b) The specific picture in the red box from (a).

From Figure 4, we can observe that the F_1 score increases and then decreases with the increase of Th_1 . When Th_1 is 0.25, F_1 reaches the highest, which is 0.942. Thus, the Th_1 is fixed to 0.25 in the remaining simulations and real data experiments.

As for the threshold of angle merging Th_2 , since the high-resolution DOA estimation is not considered, it is supposed to be lower than δ_1 . Therefore, Th_2 is set to 2° in the simulation and real data experiments.

4.4. Simulation Results

4.4.1. Sources in the Same SNR Scenarios

Two simulations are conducted under adverse conditions to assess the performance of DNN-SW, including low SNR and array imperfections. In this part, the high dynamic condition is not considered, which means that the SNR of all the sources is the same.

Since the DNN-SW can estimate the number and directions of sources simultaneously, in this simulation, the two values are unknown, and both need to be obtained. To further evaluate the proposed method, the comparison methods are applied. The existing methods can rarely achieve the two tasks at the same time. Therefore, the whole task is divided

into the source number estimation part and the DOA estimation part for the comparison methods. In the source number estimation part, two conventional methods are employed for comparison, including AIC and MDL [27–31]. Then, in the DOA estimation part, FFT, MUSIC, and ESPRIT are utilized. Furthermore, since MUSIC and ESPRIT are based on the covariance matrix, the space smoothing algorithm is applied to generate the covariance matrix before estimation [32].

Firstly, we consider two sources in the low SNR situation, and both targets have the same SNR in the testing sample. Two sources impinge on this array from the directions of -6.25° and 3.18° . The SNR varies from 0 dB to 20 dB with the step of 2 dB. For each SNR, the RMSE is obtained by averaging the results of 1000 Monte Carlo (MC) runs. The source number estimation and DOA estimation results are shown in Figure 5, respectively. In (a), we can observe that the Acc of three source number estimation methods all reach very high. For MDL and DNN-SW, when SNR is larger than 4 dB, the Acc is constantly above 99.5%. For AIC, the Acc is a little lower, which is about 98%, while it is more robust to SNR. The results indicate that the proposed method can achieve advanced performance for source number estimation. Furthermore, considering AIC is more robust and the difference of RMSE based on the two methods is small due to the high Acc, and adequate MC runs, the results of AIC are used to accomplish DOA estimation for FFT, MUSIC, and ESPRIT. As for the DOA estimation results in (b), the results show that as the SNR increases, the RMSE of all methods decreases. Among all the methods, DNN-SW performs best when SNR is below 18 dB. The results indicate that compared with other methods, DNN-SW can achieve better DOA estimation performance in low SNR conditions due to its strong data-fitting capability.

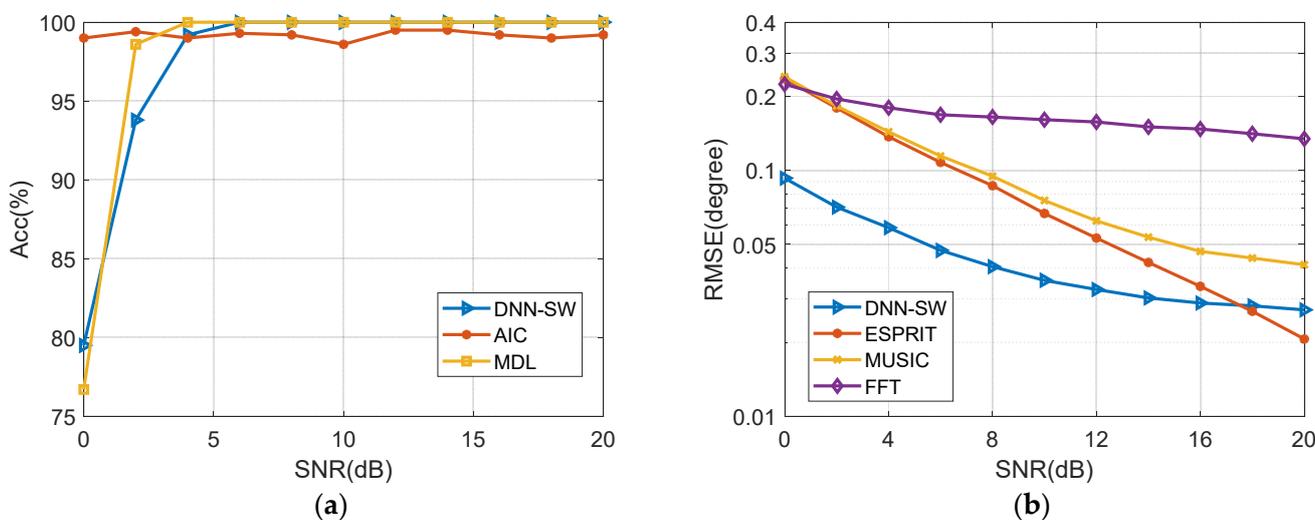


Figure 5. The DOA estimation performance with different SNRs. (a) Acc (b) RMSE.

Moreover, three kinds of array imperfections are considered, including gain inconsistent, phase inconsistent, and position perturbation [14]. It is assumed that the gain-phase inconsistency and position perturbation of different antennas are uniformly distributed within 3η dB, $30\eta^\circ$, and $0.15\lambda\eta$, respectively. η is an imperfect factor to measure the imperfect effect, and it varies from 0 to 1 with the step of 0.1. In this simulation, two sources impinge on this array from the directions of -6.25° and 3.18° . The SNR of these sources is set to 15 dB. For each η , the RMSE is obtained by averaging the results of 1000 MC runs. The results are shown in Figure 6. From (a), it can be observed that the three methods can all estimate the source number accurately. The Acc of DNN-SW is above 99% when η is 0.8 and can reach 92% even η is 1, which means the source number estimation results of DNN-SW are reliable. For MDL and AIC, the results are both satisfying, and they show a similar law to the previous simulation. In this case, the number estimation results of AIC

are still considered the basis for DOA estimation comparison methods. (b) gives the RMSE of different methods, and DNN-SW consistently performs better than other methods when the error increases. The results show that the deep learning method is more robust for array imperfections because it has the capability of adaptively learning detailed information from the input data.

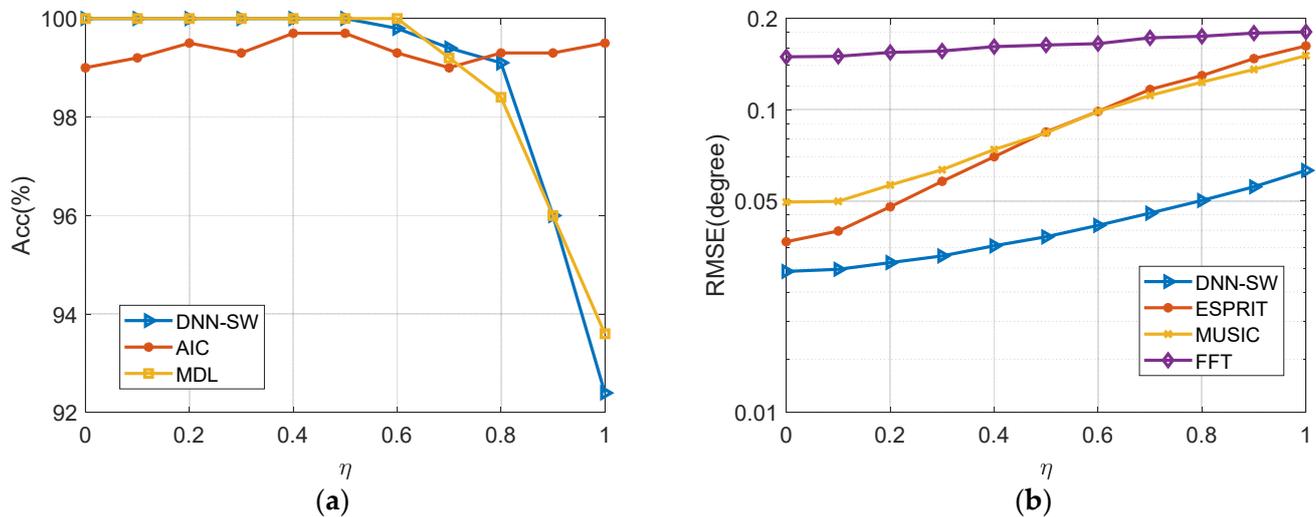


Figure 6. The DOA estimation performance with array imperfection. η is an imperfect factor. The gain-phase error and position perturbation are 3η dB, $30\eta^\circ$, and $0.15\lambda\eta$, respectively. (a) Acc (b) RMSE.

4.4.2. Sources in High Dynamic SNR Scenarios

In this part, we focus on the high dynamic SNR scenarios, which means the SNRs of different sources are different. Since the results in Section 4.4.1 show that the deep learning methods perform better under adverse conditions than conventional methods, in this simulation, two deep learning methods for DOA estimation are applied as comparisons, including DNN-NSW and Deep convolution network (DCN) [11]. The difference between DNN-NSW and DNN-SW is that there is no overlapping part between adjacent sub-regions in DNN-NSW. That is to say, in DNN-NSW, the step size of the sliding window δ_2 is set to 3° , which is the same as the δ_1 . So, a source will only appear in one sub-region. The remaining parameters and configurations of the two methods are the same. Furthermore, due to DCN methods needing multiple snapshots, the number of snapshots is set to 50.

Firstly, we change the directions of the two sources to assess the DOA estimation performance. The first source θ_1 varies from -49.59° to 49.41° with the step of 1° , and the direction of the second source θ_2 is set to $\theta_1 + 6.25^\circ$. The directions of the sources are all off-grid. When the SNR of the two sources is both 10 dB, the estimation results of DCN and DNN-SW are shown in Figure 7a, b, and c, respectively. We can observe that the three deep learning methods can achieve satisfying performance in the ideal case. Figure 7d–f depicts the results when the SNRs of two sources are 10 dB and 18 dB, respectively. The results demonstrate that Δ SNR between two sources severely degrades the performance of DCN; the source with lower SNR is rarely estimated. By contrast, DNN-SW can significantly alleviate the problem due to the design of the sub-region network structure. The lower SNR source is estimated in its sub-region, and the influence of the other sources can be reduced. In addition, compared with DNN-NSW, we can infer that the sliding window can improve the accuracy of estimating the number of sources and the performance of DOA estimation.

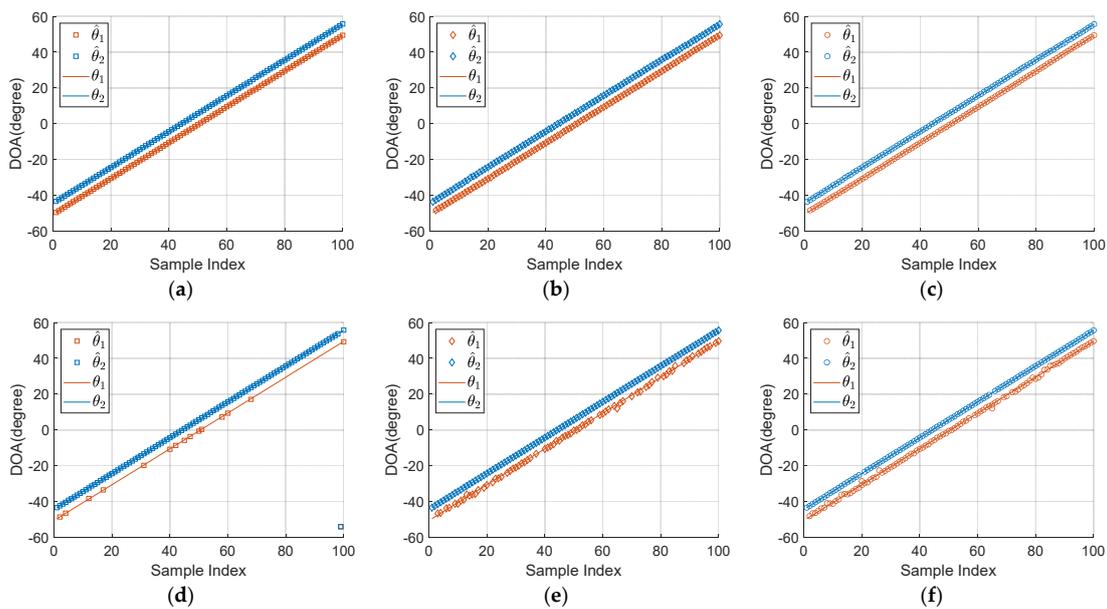


Figure 7. The DOA estimation of off-grid sources. First row: the SNR of the two sources is 15 dB. Second row: the SNR of two sources is 10 dB and 18 dB. (a) DCN, Acc:100%, RMSE:0.061, (b) DNN-NSW, Acc:100%, RMSE:0.149 (c) DNN-SW, Acc:100%, RMSE:0.128 (d) DCN, Acc:15%, RMSE:0.129, (e) DNN-NSW, Acc:78%, RMSE:0.396 (f) DNN-SW, Acc:97%, RMSE:0.262.

Additionally, to investigate the effect of the difference of SNRs between two sources, the direction of the two sources is fixed at 6.28° and 15.72° , and the SNR of the sources are configured as 10 dB and $10 \text{ dB} + \Delta\text{SNR}$, respectively. In the simulation, ΔSNR varies from 0 dB to 10 dB, and the RMSE for each ΔSNR is obtained by calculating the average results of 1000 MC runs. As shown in Figure 7d, since DCN may miss the low SNR source, the Acc of three methods is also discussed. The results are shown in Figure 8a; it can be seen that DNN-SW can precisely estimate the source number even for high ΔSNR , while DCN fails to estimate the source number when $\Delta\text{SNR} > 5$ dB. As for the RMSE given in Figure 8b, it can be observed that the RMSE of DNN-SW is much lower than DNN-NSW due to the overlapping design and repeat estimation of the sources. As for DCN, it achieves a more precise estimation when ΔSNR is small because the input contains information from multiple snapshots. However, with the increase of ΔSNR , DNN-SW shows its advantages due to the design of the sub-region network structure.

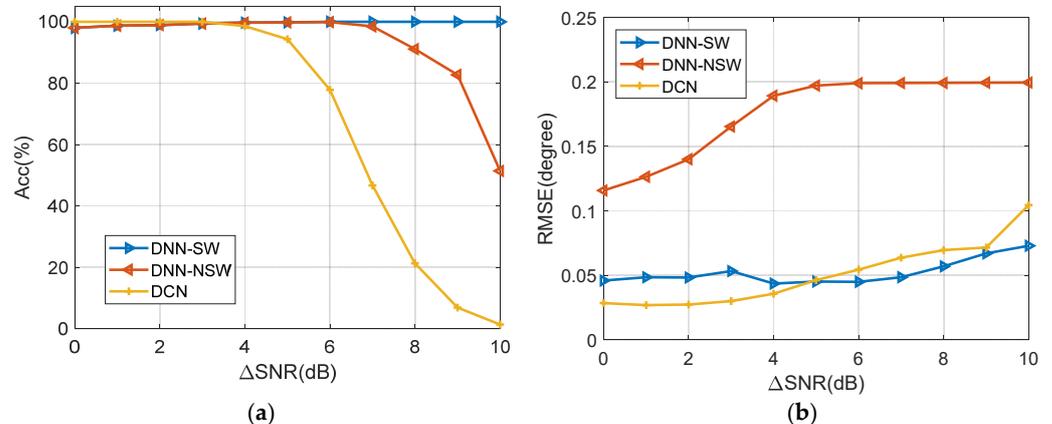


Figure 8. The DOA estimation performance of two unequally-powered sources from the directions of 6.28° and 15.72° . (a) Acc, (b) RMSE.

4.5. Real Data Experiment Results

To further evaluate the practical application value of DNN-SW, the real data are collected using MMWCAS-RF-EVM radar in the practical scenario, and the experiments are conducted based on the real data. The specific configurations of the radar antennas are described Figure 9. It has 12-transmit and 16-receive antennas, resulting in 86 non-overlapping azimuth virtual arrays. In this experiment, there are 40 virtual arrays considered. The data collection scenario is shown in Figure 10. Two different corner reflectors are fixed at a distance of 6 m from the radar, and their directions relative to the radar are -7.2° and 4.8° , respectively. In [13], the RMSE of ESPRIT is lower than 0.01 when the array element is 16, the number of snapshots is 1000, and the SNR of sources is 15 dB. Based on this result, in our experimental condition, the ground-truth directions of the corner reflectors are calculated using 86 virtual arrays and 1000 snapshots by the ESPRIT methods, and the RMSE will be lower than 0.01. Therefore, it is considered the ground truth in this real data experiment.

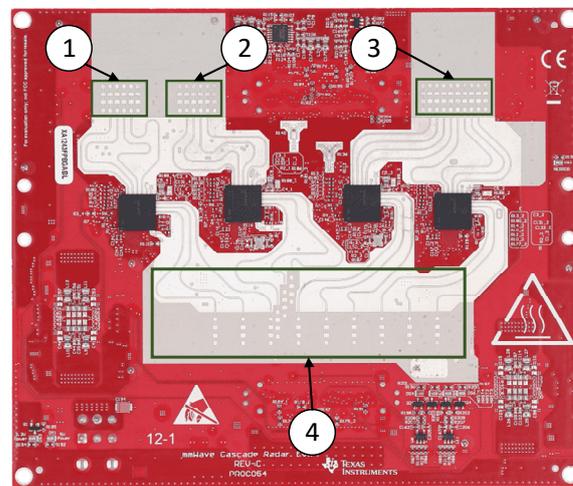
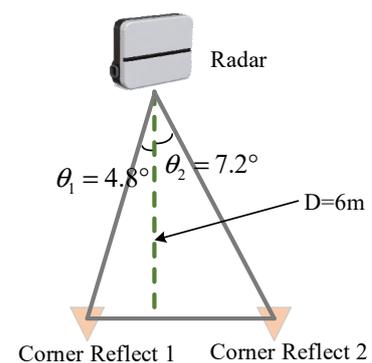


Figure 9. The PCB antenna arrays of MMWCAS-RF-EVM radar. 1, 2, 3 are the receiving antennas and 4 is the transmitting antennas.



(a)



(b)

Figure 10. The measured scenario with 77 GHz millimeter-wave array. (a) The measured scenario, (b) The scenario schematic.

Based on the measured reflected signal, the DOA estimation results of four methods are shown in Figure 11 and Table 3. We can observe from the spectrum in Figure 11 that the difference in the SNR of the two corner reflectors is 5.3 dB. In this case, DNN-SW performs best whose RMSE is only 0.071. For the conventional methods, the estimation error is larger, and MUSIC performs best among them. As for the deep learning method, DCN, the higher

source is estimated more accurately compared with most conventional methods, while the weaker source is missed. The results verify the effectiveness of DNN-SW in the practical application and indicate that the structure of our method can improve performance in high dynamic scenarios.

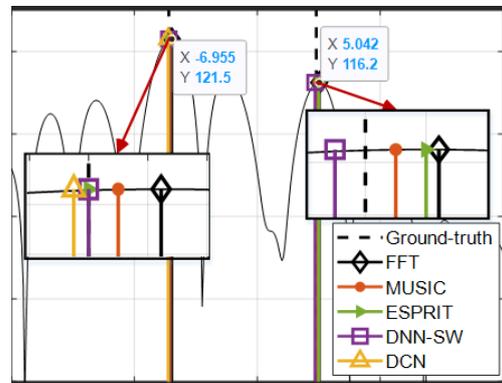


Figure 11. The DOA estimation of two different corner reflectors.

Table 3. DOA estimation results of the real data.

	Target 1 (−7.2°)		Target 2 (4.8°)		RMSE
	Estimated Result	Estimated Error	Estimated Result	Estimated Error	
FFT	−6.955°	0.245°	5.042°	0.242°	0.243
MUSIC	−7.1°	0.1°	4.9°	0.1°	0.1
ESPRIT	−7.2°	0°	5.0°	0.2°	0.141
DCN ¹	−7.25°	0.05°	/	/	/
DNN-SW ²	−7.2°	0°	4.7°	0.1°	0.071

¹ Deep convolution network (DCN) was proposed in [11]. ² Deep neural network with sliding window (DNN-SW) is proposed in this paper.

5. Conclusions

In this paper, a deep neural network framework with the angular sliding window is proposed for DOA estimation in highly dynamic scenarios. This method divides FOV into a set of sub-regions. In each sub-region, the sources are separately estimated. A detector network and an estimator network are designed for source detection and estimation. Based on the assumption that there is at most one source in each sub-region, each network can be trained with single-source data, which alleviates the requirement of training data and improves DOA estimation performance in highly dynamic scenarios. Simulation results verify the effectiveness of DNN-SW, and the results show that it can significantly estimate the source direction in highly dynamic SNR scenarios. Furthermore, the experiment results on real data show that the RMSE of the proposed method is 0.071, which is superior to FFT, MUSIC, ESPRIT, and DCN.

Author Contributions: Conceptualization, Y.L. and Y.W.; methodology, Y.L., Y.W. and Z.H.; software, Z.H. and C.L.; validation, Z.H., Y.L. and C.L.; formal analysis, C.L. and L.Z.; investigation, Y.L. and Z.H.; resources, Y.W. and L.Z.; data curation, J.W., Y.Z. and H.L.; writing—original draft preparation, Y.L. and Z.H.; writing—review and editing, Y.W., C.L., L.Z., J.W., Y.Z. and H.L.; visualization, J.W., Y.Z. and H.L.; supervision, Y.W. and L.Z.; project administration, Y.W., Y.L. and Z.H.; funding acquisition, Y.W. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China (Grant No. 2018YFE0202101, Grant No. 2018YFE0202103), China Postdoctoral Science Foundation (Grant No. 2021M690412), Natural Science Foundation of Chongqing, China (Grant No. cstc2020jcyj-msxmX0812) and project ZR2021MF134 supported by Shandong Provincial Natural Science Foundation.

Data Availability Statement: Data available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, Q.; Hu, X.; Deng, X.; Buris, N.E. DoA Estimation Using Neural Tangent Kernel under Electromagnetic Mutual Coupling. *Electronics* **2021**, *10*, 1057. [\[CrossRef\]](#)
2. Hassanien, A.; Amin, M.G.; Zhang, Y.D.; Ahmad, F. High-resolution single-snapshot DOA estimation in MIMO radar with colocated antennas. In Proceedings of the 2015 IEEE Radar Conference (RadarCon), Arlington, VA, USA, 10–15 May 2015; pp. 1134–1138.
3. Xiang, H.; Chen, B.; Yang, T.; Liu, D. Improved De-Multipath Neural Network Models With Self-Paced Feature-to-Feature Learning for DOA Estimation in Multipath Environment. *IEEE Trans. Veh. Technol.* **2020**, *69*, 5068–5078. [\[CrossRef\]](#)
4. Chen, H.; Chen, K.; Cheng, K.; Chen, Q.; Fu, Y.; Li, L. An Efficient Hardware Accelerator for the MUSIC Algorithm. *Electronics* **2019**, *8*, 511. [\[CrossRef\]](#)
5. Jung, Y.; Jeon, H.; Lee, S.; Jung, Y. Scalable ESPRIT Processor for Direction-of-Arrival Estimation of Frequency Modulated Continuous Wave Radar. *Electronics* **2021**, *10*, 695. [\[CrossRef\]](#)
6. Sun, S.; Petropulu, A.P.; Poor, H.V. MIMO Radar for Advanced Driver-Assistance Systems and Autonomous Driving: Advantages and Challenges. *IEEE Signal Process. Mag.* **2020**, *37*, 98–117. [\[CrossRef\]](#)
7. Wei, R.; Wang, Q.; Zhao, Z. Two-Dimensional DOA Estimation Based on Separable Observation Model Utilizing Weighted L1-Norm Penalty and Bayesian Compressive Sensing Strategy. In Proceedings of the 2017 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, China, 21–23 July 2017; pp. 1764–1768.
8. Bosse, J.; Rabaste, O. Subspace Rejection for Matching Pursuit in the Presence of Unresolved Targets. *IEEE Trans. Signal Process.* **2018**, *66*, 1997–2010. [\[CrossRef\]](#)
9. Wei, Z.; Wang, W.; Dong, F.; Liu, Q. Gridless One-Bit Direction-of-Arrival Estimation Via Atomic Norm Denoising. *IEEE Commun. Lett.* **2020**, *24*, 2177–2181. [\[CrossRef\]](#)
10. Barthelme, A.; Utschick, W. A Machine Learning Approach to DoA Estimation and Model Order Selection for Antenna Arrays With Subarray Sampling. *IEEE Trans. Signal Process.* **2021**, *69*, 3075–3087. [\[CrossRef\]](#)
11. Wu, L.; Liu, Z.-M.; Huang, Z.-T. Deep Convolution Network for Direction of Arrival Estimation With Sparse Prior. *IEEE Signal Process. Lett.* **2019**, *26*, 1688–1692. [\[CrossRef\]](#)
12. Guo, Y.; Zhang, Z.; Huang, Y.; Zhang, P. DOA Estimation Method Based on Cascaded Neural Network for Two Closely Spaced Sources. *IEEE Signal Process. Lett.* **2020**, *27*, 570–574. [\[CrossRef\]](#)
13. Papageorgiou, G.; Sellathurai, M.; Eldar, Y. Deep Networks for Direction-of-Arrival Estimation in Low SNR. *IEEE Trans. Signal Process.* **2021**, *69*, 3714–3729. [\[CrossRef\]](#)
14. Liu, Z.-M.; Zhang, C.; Yu, P.S. Direction-of-Arrival Estimation Based on Deep Neural Networks With Robustness to Array Imperfections. *IEEE Trans. Antennas Propagat.* **2018**, *66*, 7315–7327. [\[CrossRef\]](#)
15. Chen, P.; Chen, Z.; Liu, L.; Chen, Y.; Wang, X. SDOAnet: An Efficient Deep Learning-Based DOA Estimation Network for Imperfect Array. *arXiv* **2022**, arXiv:2203.10231.
16. Wu, X.; Yang, X.; Jia, X.; Tian, F. A Gridless DOA Estimation Method Based on Convolutional Neural Network With Toeplitz Prior. *IEEE Signal Process. Lett.* **2022**, *29*, 1247–1251. [\[CrossRef\]](#)
17. Su, X.; Hu, P.; Liu, Z.; Shi, J.; Li, X. Deep Alternating Projection Networks for Gridless DOA Estimation With Nested Array. *IEEE Signal Process. Lett.* **2022**, *29*, 1589–1593. [\[CrossRef\]](#)
18. Häcker, P.; Yang, B. Single snapshot DOA estimation. *Adv. Radio Sci.* **2010**, *8*, 251–256. [\[CrossRef\]](#)
19. Zhang, L.; Shi, C.; Niu, J.; Ji, Y.; Jonathan Wu, Q.M. DOA Estimation for HFSWR Target Based on PSO-ELM. *IEEE Geosci. Remote Sensing Lett.* **2022**, *19*, 1–5. [\[CrossRef\]](#)
20. Richards, M.A. *Fundamentals of Radar Signal Processing*, 2nd ed.; McGraw-Hill Education: New York, NY, USA, 2014; ISBN 978-0-07-179833-4.
21. Yang, Y.; Gao, F.; Qian, C.; Liao, G. Model-Aided Deep Neural Network for Source Number Detection. *IEEE Signal Process. Lett.* **2020**, *27*, 91–95. [\[CrossRef\]](#)
22. Lima de Oliveira, M.L.; Bekooij, M.J.G. ResNet Applied for a Single-Snapshot DOA Estimation. In Proceedings of the 2022 IEEE Radar Conference (RadarConf22), New York, NY, USA, 21–25 March 2022; pp. 1–6.
23. Huang, H.; Xu, H.; Wang, X.; Silamu, W. Maximum F1-Score Discriminative Training Criterion for Automatic Mispronunciation Detection. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 787–797. [\[CrossRef\]](#)
24. Pillai, I.; Fumera, G.; Roli, F. Designing multi-label classifiers that maximize F measures: State of the art. *Pattern Recognit.* **2017**, *61*, 394–404. [\[CrossRef\]](#)
25. Calders, T.; Esposito, F.; Hüllermeier, E.; Meo, R. (Eds.) *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, 15–19 September 2014. Proceedings, Part II*; Lecture Notes in Computer Science; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; Volume 8725, ISBN 978-3-662-44850-2.
26. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [\[CrossRef\]](#)

27. DeRidder, F.; Pintelon, R.; Schoukens, J.; Gillikin, D.P. Modified AIC and MDL Model Selection Criteria for Short Data Records. *IEEE Trans. Instrum. Meas.* **2005**, *54*, 144–150. [[CrossRef](#)]
28. Ding, J.; Tarokh, V.; Yang, Y. Bridging AIC and BIC: A New Criterion for Autoregression. *IEEE Trans. Inform. Theory* **2018**, *64*, 4024–4043. [[CrossRef](#)]
29. Seghouane, A.-K. Asymptotic bootstrap corrections of AIC for linear regression models. *Signal Process.* **2010**, *90*, 217–224. [[CrossRef](#)]
30. Huang, L.; So, H.C. Source Enumeration Via MDL Criterion Based on Linear Shrinkage Estimation of Noise Subspace Covariance Matrix. *IEEE Trans. Signal Process.* **2013**, *61*, 4806–4821. [[CrossRef](#)]
31. Bazzi, A.; Slock, D.T.M.; Meilhac, L. Detection of the number of superimposed signals using modified MDL criterion: A random matrix approach. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 4593–4597.
32. Liao, W.; Fannjiang, A. MUSIC for single-snapshot spectral estimation: Stability and super-resolution. *Appl. Comput. Harmon. Anal.* **2016**, *40*, 33–67. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.